Implementing the Bayesian Paradigm: Reporting Research Results over the World-Wide Web

Harold P. Lehmann[1], MD PhD, Martin R. Wachter, MES
Johns Hopkins School of Medicine, University of Baltimore

For decades, statisticians, philosophers, medical investigators and others interested in data analysis have argued that the Bayesian paradigm is the proper approach for reporting the results of scientific analyses for use by clients and readers. To date, the methods have been too complicated for non-statisticians to use. In this paper we argue that the World-Wide Web provides the perfect environment to put the Bayesian paradigm into practice: the likelihood function of the data is parsimoniously represented on the server side, the reader uses the client to represent her prior belief, and a downloaded program (a Java applet) performs the combination. In our approach, a different applet can be used for each likelihood function, prior belief can be assessed graphically, and calculation results can be reported in a variety of ways. We present a prototype implementation, `BayesApplet`*, for two-arm clinical trials with normally-distributed outcomes, a prominent model for clinical trials. The primary implication of this work is that publishing medical research results on the Web can take a form beyond or different from that currently used on paper, and can have a profound impact on the publication and use of research results.*

INTRODUCTION

The Bayesian paradigm of using scientific research results holds that the decision maker should combine data reported in a study with her prior belief, to arrive at a posterior belief that she then uses for decision making.[1, 2] Proponents of this view have argued that this paradigm matches the thought and actions of decision makers in general, and physicians in particular.[3-5] For instance, if my prior belief represents my best estimate for the risk a particular patient has for a disease, research data can combine with that belief to help me estimate the implications of published research results for that patient.

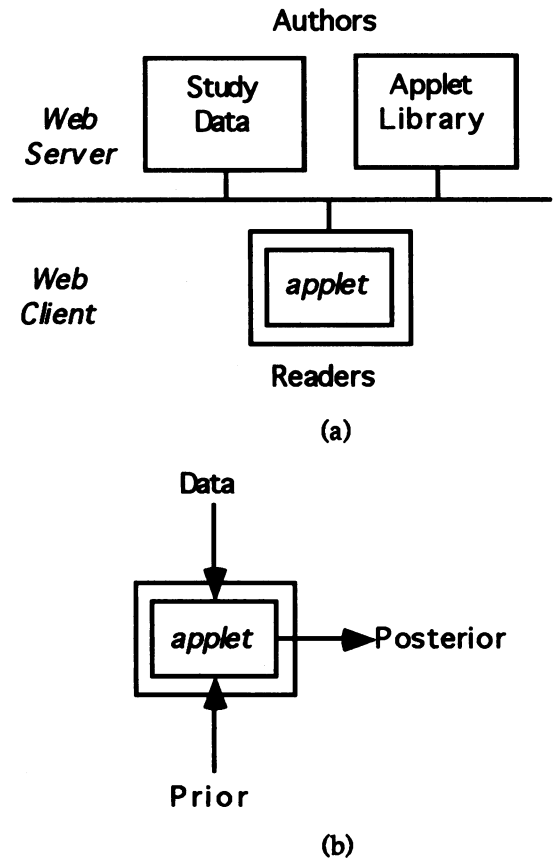

Figure 1. System architecture. (a) Client-server design (b) Specifics of applet-based interaction

Until now, methods suggested for implementing this reporting paradigm have been paper based: The investigators produce a number of graphs where a reader can locate her prior belief on one axis and find her posterior belief on another.[6] These methods have been limited in that use of the graphs require training that most non-statisticians do not have and in that publishers are reluctant to commit so much "real estate" for the depiction of these graphs. (A textual, classical-paradigm-based P value takes up much less

[1] lehmann@welchgate.welch.jhu.edu

space!) The lack of availability of proper reporting tools is one reason for the opposition against the use of Bayesian methods.[7]

With the advent of full-text publishing on the Internet, medical informaticians have the opportunity to suggest novel ways of making research data available to readers. the power of microcomputers could be brought to bear on performing the Bayesian calculation that have heretofore been problematic. In this paper we integrate these three themes: Bayesian statistics, full-text on-line publication, and the World-Wide Web.

METHODS

In Bayesian statistical analysis, the focal statistic is the likelihood function, P(specific, obtained data | any possible value of the parameter of interest). Whereas the *P* value of classical statistics produces single numbers, the likelihood function is a curve: for each potential value of a parameter (e.g., population mean systolic blood pressure), the likelihood function indicates how likely that value is, given the data. The Bayesian combination of the likelihood function with prior belief (prior P(parameter)) produces the posterior belief distribution, P(parameter | data), which is then used to produce measures of interest, such as Bayesian confidence intervals and such as optimal actionas.

Architecture

Figure 1 depicts our proposed system architecture. In this architecture, authors publish their data via Web servers. As part of the "article," the authors present the data and a link to the appropriate downloadable program (applet). Readers view those data through client Web browsers. When the data are accessed, the interactive applet is downloaded along with the data. Ideally, the applets would be part of an extensive applet library, available to all authors (and readers) on the Web. In figure 1b, the details of the client-side interaction are shown: the reader uses the applet to input her prior belief, and the applet computes the posterior belief from that prior belief and from the likelihood function.

Likelihood function

Behind each classical statistical test lies a statistical model embodying a variety of assumptions about the real world, about the experimental process, and about the quality of the data. Similarly, the likelihood function is a product of a statistical model. The essential components of a statistical model are the parametric family (normal, binomial, etc.), the parameters (mean and standard deviation for the normal distribution; success probability and number of trials, for the binomial), and the sufficient statistics—the appropriate summarization of the data for the parameters involved (e.g., sample size, mean, and standard deviation, for the mean of a normal distribution). A number of general methods have been proposed over the last decade for the computer-based representation of the likelihood function.[8, 9]* This general approach would entail all authors invoking the same Bayesian applet, but having to represent the likelihood function in some standard fashion.

In our approach, we suggest a simpler method: each applet has associated with it a canonical likelihood function. Therefore, the author chooses the appropriate applet for his data, and needs to put only the data onto his Web document (in an appropriate format; see Results); the applet produces the computable form of the likelihood function internally upon execution.

Prior belief

Assessment of prior belief can sometimes be difficult primarily because users are unfamiliar with the task. Statisticians have suggested a number of assessment methods.[10, 11] The two major approaches are to assess the belief directly and numerically, or to assume a parametric family for the prior belief, and then to assess the decision maker's estimates of the parameters. As an example of the latter approach in the case of prior belief in a mean blood pressure, the assessor would ask the decision maker if her belief would look like a normal distribution, and then would ask her to state what she thinks the mean is, a priori, and to state a standard deviation that represents her uncertainty in that estimate.

An advantage of an applet-based environment is that graphical methods can be used that have heretofore not been possible. In our prototype, we have the user graphically move the entire prior distribution curve, rather than type in values for the prior mean and standard deviation (see Results).

Posterior belief

The heart of Bayesian statistical reporting is the calculation of posterior belief. A number of general approaches are available from the Bayesian community [12, 13] However, these general methods are time intensive, because, in the general case, the Bayesian calculation requires multiple numerical

---

* The evolving file format for belief networks, a superset of statistical models, is found in http://www.research.microsoft.com/research/dtg/bnformat/default.htm

integration or simulation. A time-saving approach is to make some strong assumptions about the character of the prior distribution vis á vis the likelihood function and thereby to simplify the calculation. For instance, if the likelihood function were normal and the prior belief on the mean were normal (and the standard deviation of the likelihood function were assumed to be known), then non-integrating algorithms would be used.[14] For small models, there are direct calculations that can be performed very quickly.[15]

Associated measures

Given a posterior distribution, Bayesian analysts [15, 16] have recommended a variety of measures to aid in inference and in decision making. One that we present in this paper is the *tail probability*, P(parameter > specified value | data), that has the semantics that most readers assign to *P* values.

RESULTS

For our prototype, we chose the problem of reporting data from a two-arm randomized clinical trial where the outcomes are normally distributed. In classical statistics, a *t* test would be calculated from the data, a *P* value would be generated from the test statistic (as would a confidence interval), and the *P* value would be compared with an error limit of, say, 0.05. If the *P* value were less than 0.05, the null hypothesis would be rejected that the two treatments (assigned to each arm) result in the same outcome.

To publish the data with our Bayesian applet,* the author includes the following HTML code in his document (assuming he had access to the `bayesApplet` Java code).

```
1. <applet code=bayesApplet.class
width=600 height=400>
2. <param name=imgIdle value=
"openHand.gif">
3. <param name=imgActive value=
"closeHand.gif">
4. <param name=datacurve1 value=
"Experimental,.1,10,30">
5. <param name=datacurve2 value=
"Plecebo,.5,10,35">
6. </applet>
```

Line 1 opens the reference to the applet and reserves some screen real estate. Lines 2 and 3 tell the applet where to find two image files the applet needs for user input of prior belief. Lines 4 and 5 are the core of the data publishing: each informs the applet of the name of one arm of treatment, as well as the sufficient statistics for the results of that arm (mean, standard deviation, and sample size). Line 6 closes the reference.

In terms of our design specifications, note that the likelihood function is implied in the choice of applet (bayesAppletclass). Furthermore, the data are reported as sufficient statistics in a format determined by the applet. If the author wanted the reader to see the complete data in some tabular format, for instance, that table would have to appear elsewhere in the Web document.

To give a sense of the user's interaction with the applet, Figure 2 provides a screen shot. Data (sufficient statistics) from the server are displayed in the Input Panel and are used to create the likelihood function. The user creates the prior distribution either by typing into the Input Panel or by manipulating the graph on the Canvas. The posterior distribution is created on the fly; every time the prior distribution changes, the applet recalculates, and displays, the posterior distribution. The associated measures (in this case, only the tail probability) are calculated as well. The interaction is fast; as the user drags the prior distribution around the Canvas, the posterior curve changes immediately.

We use a novel approach to the graphical input of prior belief. The user can "grab" the prior distribution curve directly. The applet interprets horizontal movement of the curve as a change in the prior mean. It interprets vertical movement as a change in standard deviation: "up" mean greater certainty, or a smaller standard deviation; "down" means less certainty, or a larger standard deviation. A flat prior curve means that the user has no prior belief in the value of the mean of interest, called *non-informative* in the Bayesian literature.[17] Usually, a non-informative prior leads to Bayesian calculations that are numerically the same as some classical measures, but retain the more intuitive, Bayesian interpretations.

Thus, a user might typically interact with the system as follows: First, she could use a non-informative prior to see what the classical-statistical conclusion would be. Then, she could use a "real" prior, representing her true prior knowledge, to see the implications of the data for her current state of knowledge. Finally, she could use extreme priors to see how skeptical she would have to be to change her mind. This last use is the most powerful, because the interaction allows the user to argue with the data, a rhetorical stance not possible with traditional paper-based publication

---

* See http://infonet.welch.jhu.edu/~mrw//Bayes/index .html for the most current public version.

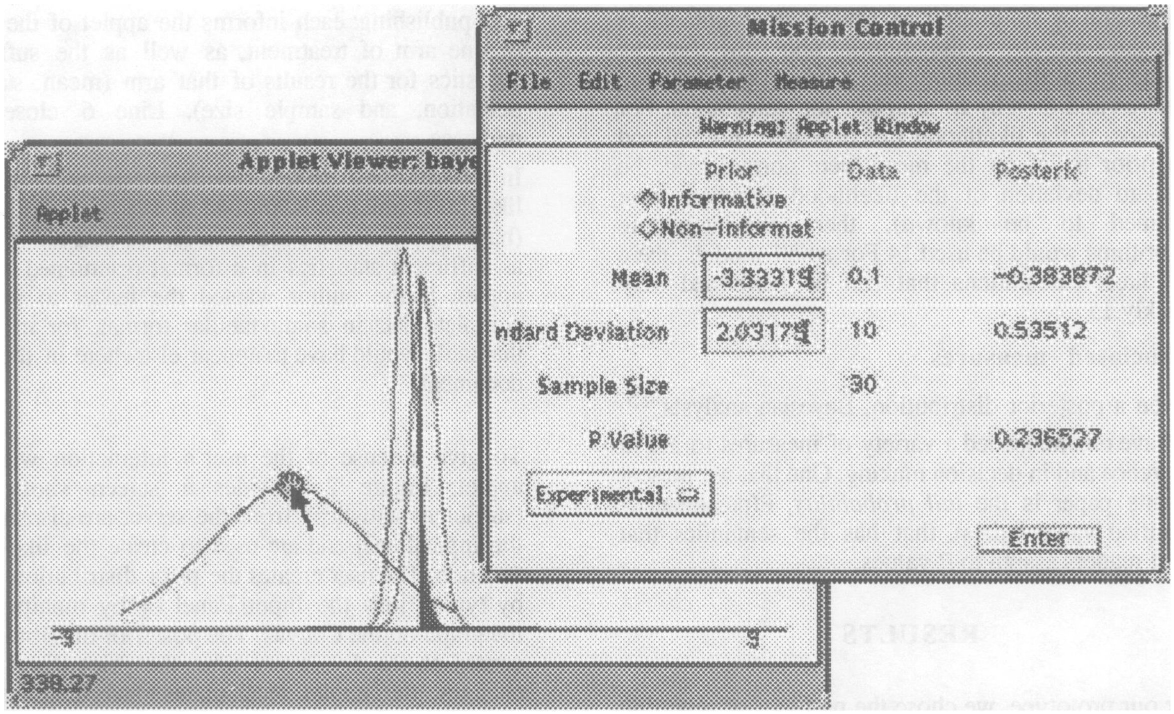

Figure 2. Screen shot of `bayesApplet`. In this example, the two treatment arms are Experimental and Placebo. The data for arm Experimental (mean of 0.1, standard deviation (sd) of 10, and sample size of 30 (hence, a standard error of 1.8)) are displayed in the "data" column of the `InputPanel`. The informative prior belief input by the user has a mean of –3.33 and sd of 2.03. The calculated posterior mean is –0.38 (a weighted average of the prior and of the data means) and the posterior sd is 0.535 (i.e., more certain than either the prior or the data). The curves for the prior belief, the posterior belief, and the likelihood function are laid out on the Canvas from left to right (color-coded). The user can alter her prior belief either by typing in the `InputPanel`, or by "grabbing" and moving the hand icon by clicking and dragging the mouse. The area under the posterior belief curve to the right of the y-axis represents the tail probability measure. The value of this Bayesian *P* value (0.24) is displayed in the `InputPanel`.

The `bayesApplet` applet itself comprises a number of objects whose classes, inheritance, and ownership relationships are shown in Figure 3. For instance, an instance of `graphCanvas` class has, as members, one instance each of a `normalPriorCurve`, a `normalPosteriorCurve`, and a `normalDataCurve`. Each of these classes are specializations of the `bayesCurve` class. An instance of `graphCanvas` displays the curves (with the possibility of the user manipulating the prior curve), while an instance of the `inputPanel` allows numerical inputs and displays numerical outputs.

The prototype discussed here was written in Java Alpha3, Release 1 of the Solaris Java Development Kit and has been tested to date on beta versions of Netscape 2.0 for Solaris and for Windows95.

DISCUSSION

The World-Wide Web is the natural environment for on-line full-text publishing: The Web can archive research results, while users can interact with the data in a Bayesian manner. The prototype presented here is only a first stab at this paradigm, but the implication is clear: The Web can be used to implement the Bayesian paradigm.

The current prototype's interface needs to be improved to be usable by physicians. Comprehension would be improved by using a wider variety of measures associated with the posterior distribution and by attaching clinically relevant interpretations to them. For instance, adding in decision thresholds would aid dramatically in this interpretation. (e.g., what difference is *clinically* significant). Furthermore, rather than simply having the user vary the prior distribution, the machine could point out at what point the conclusion has changed (sensitivity analysis).

An alternate approach to publishing research results is for the author to include all the raw data, and allow the user to use applet-based database-manipulation tools to examine the data. While such spreadsheet tools are sure to be commonly available, this form of interaction would ignore the statistical model-selection work done by the investigator, and can introduce unintended biases due to inappropriate analysis.

Computation of the posterior distribution could conceivably be performed by a server-based (cgi) program. However, that design would put the burden of computation either on the author or on the central library/server. The latter design would not scale up for the case of large numbers of readers; the former design puts too much burden on authors.

It is a commonplace to say that the Web will change the way we work and think. Yet, implementing the Bayesian paradigm on the Web truly has the potential to convert passive readers of research articles into active participants, to change the role of reader into that of decision maker, to change the rhetorical nature of publishing research results, and to fulfill the informatics goal of making the clinical research literature a real aid to clinical practice.

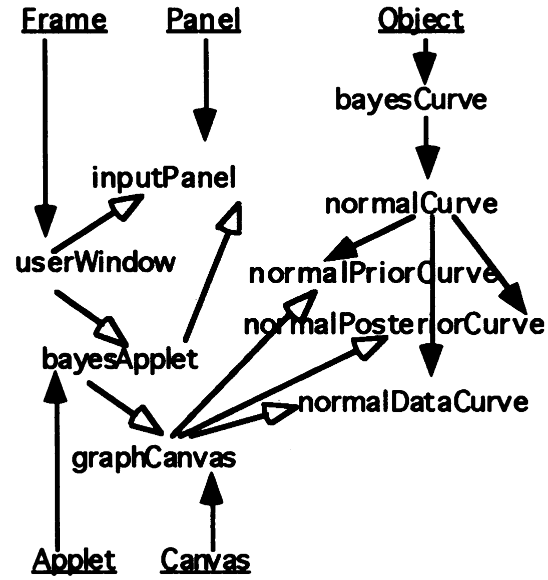

Figure 3. Bayes applet object hierarchy. Underlined names refer to Java built-in classes. Solid arrows indicate inheritance; empty-headed arrows indicates ownership by an instance.

ACKNOWLEDGMENTS

Thanks to Steve Goodman for important feedback. This research was funded by NLM grant R29-LM05647-03

REFERENCES

1. Howson C, Urbach P. Scientific reasoning: The Bayesian approach. La Salle, IL: Open Court Publishing Co., 1989.
2. Lindley DV. Bayesian statistics. Philadelphia: Society for Industrial and Applied Mathematics, 1972.
3. Cornfield J. A Bayesian test of some classical hypotheses-with applications to sequential clinical trials. Journal of the American Statistical Association 1966;61(315):577-594.
4. Hilden J. Reporting clinical trials from the viewpoint of a patient's choice of treatment. Statistics in Medicine 1987;6:745-752.
5. Lehmann HP. A Bayesian computer-based approach to the physician's use of the clinical research literature [doctoral dissertation]. Stanford University, 1991.
6. Hildreth C. Bayesian statisticians and remote clients. Econometrica 1963;31(3):422-438.
7. Efron B. Why isn't everyone a Bayesian? The American Statistician 1986;40:1-5.
8. Gilks WR, Thomas A, Spiegelhalter DJ. A language and program for complex Bayesian modelling. The Statistician 1994;43(1):169–177.
9. Lehmann HP, Shachter RD. A physician-based architecture for the construction and use of statistical models. Methods of Information in Medicine 1994;33:423–432.
10. Winkler RL. The assessment of prior distributions in Bayesian analysis. Journal of the American Statistical Association 1967;62:776-800.
11. Spiegelhalter DJ. Statistical methodology for evaluating gastrointestinal symptoms. Clinics in Gastroenterology 1985;14(3):489–515.
12. Smith AFM, Roberts GO. Bayesian computation via the Gibbs sampler and related markov chain Monte Carlo methods. Journal of the Royal Statistical Society (Series B) 1993;55(1):3–23.
13. Tanner MA. Tools for Statistical Inference. 2nd ed. New York: Springer-Verlag, 1993.
14. A linear approximation method for probabilistic inference. Proceedings of the Fourth Annual Workshop on Uncertainty in Artificial Intelligence; 1988; University of Minnesota, August 19-20.
15. Box GEP, Tiao GC. Bayesian inference in statistical analysis: Addison-Wesley Publishing Company, 1973.
16. Berger JO. Statistical decision theory and Bayesian analysis. Second ed. New York: Springer-Verlag, 1985.
17. Bernardo JM. Reference posterior distributions for Bayesian inference. Journal of the Royal Statistical Society, Series B 1979;41(2):113-147.